PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification [4]: C07H 21/00, C12N 15/00, 9/08 C12N 5/00, 1/20, 1/16

A1

(11) International Publication Number: WO 88/ 07541

(43) International Publication Date: 6 October 1988 (06.10.88)

(21) International Application Number: PCT/US88/00835

(22) International Filing Date: 16 March 1988 (16.03.88)

(31) Priority Application Number: 034,017

(32) Priority Date: 1 April 1987 (01.04.87)

(33) Priority Country: US

(71) Applicant: CHIRON CORPORATION [US/US]; 4560 Horton Street, Emeryville, CA 94608 (US).

(72) Inventors: MULLENBACH, Guy ; 339 63rd Street, Oakland, CA 94618 (US). HALLEWELL, Robert, A. ; 27 Priest Street, San Francisco, CA 94109 (US).

(74) Agents: MONROY, Gladys, H. et al.; Ciotti & Murashige, Irell & Manella, 545 Middlefield Road, Suite 200, Menlo Park, CA 94025 (US).

(81) Designated States: AT (European patent), BE (European patent), CH (European patent), DE (European patent), FR (European patent), GB (European patent), IT (European patent), JP, LU (European patent), NL (European patent), SE (European patent).

Published
With international search report.

(54) Title: POLYNUCLEOTIDE CONSTRUCT ENCODING A POLYPEPTIDE SUBSTANTIALLY SIMILAR TO HUMAN GLUTATHIONE PEROXIDASE

(57) Abstract

Human glutathione peroxidase (GPx) is of potential therapeutic use for a variety of disorders stemming from a variety of peroxides which are metabolic products. However, sources of human GPx are extremely limited. The present invention provides a method for synthesizing human GPx on a large scale by use of recombinant DNA techniques. In the present invention, a polynucleotide sequence coding for human GPx is provided. The polynucleotide sequence, or its modified forms, may be inserted into expression vectors allowing the synthesis of recombinant products which include human GPx, fragments of human GPx, analogs of human GPx, and analogs of fragments of human GPx. These recombinant polypeptide products have potential therapeutic use. In addition, they may be used to raise antibodies, both monoclonal and polyclonal, to human GPx. Antibodies to human GPx are useful for the diagnosis of deficiencies in human of GPx and of selenium. Also, analog of human GPx, active fragments of GPx, and monoclonal antibodies to human GPx are all useful for determining how the enzyme interacts with its various substrates, as well as how the enzyme may be targeted for hydrophilic and hydrophobic sites. This knowledge will be useful in designing analogs of human GPx for specific therapeutic uses.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	**FR**	France	**ML**	Mali
AU	Australia	**GA**	Gabon	**MR**	Mauritania
BB	Barbados	**GB**	United Kingdom	**MW**	Malawi
BE	Belgium	**HU**	Hungary	**NL**	Netherlands
BG	Bulgaria	**IT**	Italy	**NO**	Norway
BJ	Benin	**JP**	Japan	**RO**	Romania
BR	Brazil	**KP**	Democratic People's Republic of Korea	**SD**	Sudan
CF	Central African Republic			**SE**	Sweden
CG	Congo	**KR**	Republic of Korea	**SN**	Senegal
CH	Switzerland	**LI**	Liechtenstein	**SU**	Soviet Union
CM	Cameroon	**LK**	Sri Lanka	**TD**	Chad
DE	Germany, Federal Republic of	**LU**	Luxembourg	**TG**	Togo
DK	Denmark	**MC**	Monaco	**US**	United States of America
FI	Finland	**MG**	Madagascar		

POLYNUCLEOTIDE CONSTRUCT ENCODING A POLYPEPTIDE SUBSTANTIALLY SIMILAR TO HUMAN GLUTATHIONE PEROXIDASE

Background

1. Technical Field

This invention relates to recombinant DNA, more specifically to the cloning and expression of DNA which encodes human glutathione peroxidase.

2. Background Art

Peroxides and oxygen derived free radicals, generated as a result of a variety cellular reactions, have been implicated in the generation of tissue damage involving membranes, soluble proteins, and DNA. For example, it is thought that a major contributor to reperfusion injuries associated with heart attacks, stroke, intestinal ischemia, frost bite, spinal cord trauma, and organ transplantation are oxygen-derived free radicals generated by the xanthine dehydrogenase system. Hypobaric oxygen treatment is also associated with ischemic type injuries. In addition, patients suffering from genetic or alimentary deficiencies in glutathione peroxidase (GPx) suffer from hemolytic episodes if they are exposed to drugs or xenobiotics generating superoxide radicals, H_2O_2, or lipid peroxides in the red cells. Lipid peroxidation is

thought to contribute significantly to the aging of the brain; drugs such as chlorpromazine which are extensively used in psychiatry, inhibit lipid peroxidation and induce the synthesis of glutathione peroxidase.

Glutathione peroxidase (GPx) is a selenium-containing enzyme present in mammals and birds which catalyzes the reaction

$$ROOH + 2\ GSH \longrightarrow 2\ GSSG + ROH + H_2O$$

where R may be an aliphatic or aromatic organic group, or hydrogen. This enzyme is of considerable importance because of its ability to remove a wide variety of hydroperoxides, but particularly lipid peroxides and H_2O_2.

GPx, in conjunction with its cohort enzyme superoxide dismutase (SOD), acts as a scavenger system for superoxide radicals and for peroxides. SOD converts a superoxide radical to H_2O_2, which in turn is converted to H_2O in the presence of reduced glutathione and glutathione peroxidase. It is of interest that SOD, itself, is damaged by its product, H_2O_2. Also, H_2O_2 can form a highly reactive and damaging hydroxyl radical or superoxide according to the following reactions:

$$Fe^{+3} + H_2O_2 \rightarrow Fe^{+2} + O_2^- + 2H^+$$

and

$$Fe^{+2} + H_2O_2 \rightarrow Fe^{+3} + OH\bullet + OH^-$$

This damage may be prevented by adequate levels of GPx. In sum, GPx is a key enzyme in reducing cellular damage which could occur from the generation of peroxides and superoxide radicals.

In addition GPx serves other roles. For example, it modulates the formation of leukotrienes, protaglandins, and thromboxanes; this modulation is the result of the GPx's reduction of intermediate lipid peroxides generated in the arachidonic cascade. Thus, the modulation of GPx activity could control inflammatory and hypersensitivity reactions.

It is also possible that GPx is a factor in the development of coronary heart disease. A deficiency in this enzyme could facilitate the occurrence of lesions in the endothelial lining of the vessels, it could impair the formation of the antiaggregatory prostacyclin, and it could lead to the premature aging of red cells. Gpx may also have a role in cataract formation in lens. Not only is there a correlation between age and decrease in GPx activity in human lens, it has also been shown that lenticular GPx activity is inversely proportional to severity of human cataract. Moreover, the onset of nuclear cataract of the lens is accompanied by a precipitous loss of GPx activity.

GPx has been isolated from different mammalian sources, including bovine erythrocytes, ovine erythrocytes, rat liver, human erythrocytes, and human placenta. Cf Flohe, in FREE RADICALS IN BIOLOGY, VOL V, (Pryor, W.A., ed., Academic Press, N.Y.) pages 223-227.

The enzyme isolated from bovine erythrocytes is the most thoroughly studied, and serves as a prototype for structural considerations. GPx is a soluble tetrameric protein of molecular weight about 85,000. It contains four identical subunits of approximately 21,000

daltons. An unusual component of the enzyme is selenocysteine, one residue of which occurs in each subunit, as revealed by X-ray diffraction analysis and complete sequencing of bovine enzyme. W.A. Gunzler et al (1984), Hoppe-Seyler;s Z. Physiol. Chem 364, 195. Studies with rat liver slices suggest that selenium is incorporated into the enzyme via a translational pathway. W.C. Hawkes and A.L. Tappel (1983), Biochim. Biophys. Acta 739, 225.

Human GPx could be used therapeutically to overcome the effects of genetic deficiencies in the enzyme. It also has potential therapeutic use in overcoming the effects caused by ischemic trauma, and as an adjunct to hypobaric oxygen treatment. In addition, it could find use in the mental disorders related to aging, or which are currently treated with chlorpromazine. It also may be useful for the treatment, both prophylactic and therapeutic, of cataracts. However, a limitation in the source of this enzyme, i.e. human tissue, is to be expected.

Activators and inhibitors of GPx activity may have use in modulating the inflammatory and hypersensitivity reactions which result from products produced by the arachidonic acid cascade. To efficiently design specific modulators, it is necessary to isolate human GPx in sufficient quantity so that it can be structurally characterized and its mechanisms of action understood.

Disclosure of the Invention

One embodiment of the invention is a polynucleotide construct encoding a polypeptide substantially similar to human GPx. The polynucleotide

construct contains a heterologous region, which contains a coding sequence for the polypeptide. Polypeptides substantially similar to human GPx include: human GPx; fragments of human GPx; analogs of human GPx, and analogs of fragments of human GPx.

Another embodiment of the invention is a composition comprising cDNA encoding human GPx.

Still another embodiment of the invention is a method of producing a recombinant polypeptide substantially similar to human GPx. The method for producing the polypeptide consists of several steps which include the following. First, providing a population of transformed cells containing a vector, in which the vector consists of: (1) a coding sequence for a polypeptide substantially similar to human GPx; and (2) sequences which allow the expression of the coding sequence in the cells, wherein these sequences consist of at least a promoter which controls transcription of the coding sequence, a ribosomal binding sequence, a translation termination sequence, and a transcription termination sequence. Second, growing the population of transformed cells under conditions whereby the polypeptide substantially similar to human GPx is expressed. Finally, recovering the expressed polypeptide.

Yet another embodiment of the invention is a composition containing a polypeptide synthesized by expression of the above-mentioned polynucleotide construct.

The invention also consists of a composition which contains a polypeptide synthesized by expression of the polynucleotide construct which encodes a polypeptide substantially similar to human GPx.

Brief Description of Drawings

Figure 1 shows the synthetic oligonucleotides used as probes in the isolation of cDNA encoding bovine GPx.

Figure 2 shows a comparison of bovine, mouse, and human GPx cDNA sequences.

Figure 3 shows the homology of bovine, mouse, and human preprocessed GPx putative translation products.

Modes for Carrying Out the Invention

The practice of the present invention will employ, unless otherwise indicated, conventional techniques of molecular biology, microbiology, and recombinant DNA, which are within the skill of the art. Such techniques are explained fully in the literature. See e.g., Maniatis, Fritsch & Sambrook, MOLECULAR CLONING: A LABORATORY MANUAL (1982); DNA CLONING, Volumes I and II (D.N. Glover ed. 1985); OLIGONUCLEOTIDE SYNTHESIS (M.J. Gait ed. 1984); NUCLEIC ACID HYBRIDIZATION (B.D. Hames & S.J. Higgins eds. 1984); TRANSCRIPTION AND TRANSLATION (B.D. Hames & S.J. Higgins eds. 1984); ANIMAL CELL CULTURE (R.I. Freshney ed. 1986); IMMOBILIZED CELLS AND ENZYMES (IRL Press, 1986); B. Perbal, A PRACTICAL GUIDE TO MOLECULAR CLONING (1984); and the series, METHODS IN ENZYMOLOGY (S. Colowick and N. Kaplan eds, Academic Press, Inc.).

In describing the present invention, the following terminology will be used in accordance with the definitions set out below.

A "nucleotide construct" refers to the polymeric form of nucleotides, either ribonucleotides or deoxyribonucleotides. This term refers only to the primary structure of the molecule. Thus, this term includes double- and single-stranded DNA as well as double- and single-stranded RNA.

A "replicon" is any genetic element (e.g., a plasmid, a chromosome, a virus) that behaves as an autonomous unit of polynucleotide replication within a cell; i.e. capable of replication under its own control.

A "vector" is a replicon in which another polynucleotide segment is attached, so as to bring about the replication and/or expression of the attached segment.

A "coding sequence" is a polynucleotide sequence which is transcribed and/or translated into a polypeptide when placed under the control of appropriate regulatory sequences. The boundaries of the coding sequence are determined by a translation start codon at the 5'-terminus and a translation stop codon at the 3'-terminus. A coding sequence can include, but is not limited to, prokaryotic sequences, cDNA from eukaryotic mRNA, genomic DNA sequences from eukaryotic DNA, and even synthetic DNA sequences. A polyadenylation signal and transcription termination sequence will usually be located 3' to the coding sequence.

A "promoter sequence" is a DNA regulatory region capable of binding RNA polymerase and initiating transcription of a downstream (i.e. in the 3' direction) coding sequence. In the present invention, the promoter sequence is bounded at its 3'-terminus by the translation start codon of a coding sequence and extends upstream (i.e. in the 5' direction) to include the minimum number of bases or elements necessary to initiate transcription at levels detectable above background. Within the promoter sequence will be found a transcription initiation site (conveniently determined by mapping with nuclease S1), as well as protein binding domains (consensus sequences) responsible for the binding of RNA polymerase.

A coding sequence is "under the control" of the promoter sequence in a cell when transcription of the coding sequence results from the binding of RNA polymerase to the promoter sequence; translation of the resulting mRNA then results in the polypeptide encoded within the coding sequence.

"Transformation" is the insertion of an exogenous polynucleotide into a host cell. The exogenous polynucleotide may be maintained as a plasmid, or alternatively, may be integrated within the host genome. In cases where the exogenous polynucleotide is RNA, the RNA or a portion thereof will be reverse transcribed into DNA prior to integration.

Two DNA sequences are "substantially homologous" when at least about 90%, and preferably at least about 95%, of the nucleotides match over the defined length of the DNA sequences. Sequences that are substantially homologous can be identified by hybridization experiments, wherein the hybridization is carried out under stringent conditions as defined for that particular system. See, e.g., Maniatis et al., supra; DNA CLONING, Vols. I & II, supra; NUCLEIC ACID HYBRIDIZATION, supra. Detection of non-hybridized regions may be, for example: by S1 nuclease degradation of non-hybridized segments followed by size analysis by gel electrophoresis; or by heteroduplex analysis by electron microscopy.

A "polypeptide which is substantially similar to human GPx" includes human GPx, analogs of human GPx and analogs of human GPx fragments. An "analog of human GPx" or of a "fragment of human GPx" is one in which the homology to human GPx is greater than about 90%, and preferably greater than about 95%. Analogs include e.g., polypeptides which lack the selenium moiety, or

ones in which the selenium moiety is joined to an amino acid other than cysteine, as well as molecules which have amino acid substitutions. Analogs of human GPx and of human GPx fragments give rise to antibodies which react preferentially with human GPx when compared to GPx derived from other species. Preferential reactivity of antibodies is assayed by competition immunoassay, which is well known to those skilled in the art. A "polypeptide which is substantially similar to human GPx" also includes fragments of human GPx, both active and inactive. An "active fragment" of human GPx is that portion of the polypeptide which will, with the appropriate cofactors, catalyze the reaction normally catalyzed by the native enzyme. An inactive fragment of human GPx is a peptide or polypeptide which has the sequence of a portion of the human GPx molecule. An inactive fragment of human GPx is recognized by its preferential reactivity with an antibody to human GPx.

A "clone" is a population of cells derived from a single cell. A "cell line" is a clone of a primary cell that is capable of stable growth in vitro for many generations.

A "heterologous region" of a polynucleotide construct is an identifiable segment of polynucleotide within the larger polynucleotide molecule that is not found in association with the larger molecule in nature. Thus, when the heterologous region encodes a mammalian gene, the heterologous region will be flanked by a nucleotide sequence that does not flank it in the genome of the source organism. Another example of a heterologous coding sequence is a construct where the coding sequence itself is not found in nature (e.g., a cDNA where the genomic coding sequence contains introns, or a synthetic sequence having one or more codons

different than the native gene). Allelic variations or naturally occurring mutational events do not give rise to a heterologous region of DNA.

A nucleotide construct encoding human GPx may be used in methods to produce the enzyme by recombinant methods. It may also be used in gene therapy to alleviate genetic disorders producing enzyme deficiency.

DNA encoding human GPx was isolated by a method which relied, in part, on nucleic acid hybridization. Since data on both the nucleotide sequence and the amino acid sequence of the human enzyme was lacking, a novel method was derived which relied on the known amino acid sequence of the bovine enzyme.

In this method, DNA probes were designed so as to have at least moderately high homology with two regions of the bovine enzyme. Degeneracies in the code were taken into consideration and the probes designed around them, as discussed in the Experimental section, infra. These DNA probes were then used to isolate by nucleic acid hybridization and cloning techniques, a full length cDNA encoding bovine GPx. Using the cDNA encoding bovine GPx as a probe, cDNAs encoding human GPx were isolated from a λ-cDNA library carrying EcoRI recombinant inserts derived from human kidney. The cDNAs encoding human GPx were subcloned, and sequenced. The sequence for human GPx, as derived from these clones, is presented in Fig. 2.

An outstanding feature of the DNA sequence for human GPx is the presence of a single opal translation stop codon (TGA) in frame within the coding region. This feature is also present in the DNA sequences for bovine and mouse GPx.

The amino acid sequence for human GPx was deduced from the nucleotide sequence encoding it. It

may be deduced from the amino acid sequence (shown in Fig. 3) that human GPx probably does not enter the secretory pathway and that the polypeptide is not N-glycosylated.

Although a method for preparing a DNA construct encoding human GPx based upon hybridization to a bovine cDNA probe has been provided, in the current invention the preparation of such constructs is not limited to this method. Utilizing the sequence information provided herein, other methods of preparing polynucleotide constructs encoding GPx may be devised. For example, the nucleotide sequence encoding human GPx may be synthesized utilizing automated DNA synthesis. See, e.g.., Edge (1981) Nature 292, 756; Nambair et al (1984) Science 223, 1299; Jay et al (1984) J. Biol. Chem. 259, 6311. Alternatively, oligonucleotides containing a portion of the sequence information may be synthesized; these may then be used as probes to screen human genomic DNA libraries and cDNA libraries. The basic strategies for preparing oligonucleotide probes and DNA libraries, as well as their screening by nucleic acid hybridization, are well known to those of ordinary skill in the art. See, e.g., DNA CLONING: VOL. I (D.P. Glover ed 1985); NUCLEIC ACID HYBRIDIZATION (B.D Hames & S.J. Higgins eds. 1985); OLIGONUCLEOTIDE SYNTHESIS (M.J. Gate ed. 1984); T. Maniatis et al., MOLECULAR CLONING: A LABORATORY MANUAL (1982); B. Perbal, A PRACTICAL GUIDE TO MOLECULAR CLONING (1984).

Once a sequence encoding human GPx has been prepared or isolated, it can be cloned into any suitable replicon to create a vector, and thereby be maintained in a composition which is substantially free of vectors that do not contain the human GPx gene (e.g., other clones derived from the library). Numerous cloning

vectors are known to those of skill in the art, and the selection of an appropriate cloning vector is a matter of choice. Examples of recombinant DNA vectors for cloning and host cells which they can transform include the bacteriophage λ (E. coli), pBR 322 (E. coli), pACYC 177 (E. coli), pKT 230 (gram-negative bacteria), pGV1106 (gram-negative bacteria), pLAFR1 (gram-negative bacteria), pME290 (non-E. coli gram-negative bacteria), pHV 14 (E. coli and Bacillus subtilis), pBD9 (Bacillus), pIJ61 (Streptomyces), pUC6 (Streptomyces), actinophage ΦC31 (Streptomyces, YIp5 (Saccharomyces, YCp19 (Saccharomyces, and bovine papilloma virus (mammalian cells). See generally, DNA CLONING: VOLS. I & II, supra; T. Maniatis et al., supra; B. Perbal, supra.

The polynucleotide sequence encoding human GPx is expressed by inserting the sequence into an appropriate replicon thereby creating an expression vector, and introducing the resulting expression vector into a compatible host.

In creating an expression vector the human GPx sequence is located in the vector with the appropriate control sequences, which include a promoter, a ribosomal binding site, and transcriptional and translational stop codons. The positioning and orientation of the coding sequence with respect to the control sequences is such that the coding sequence is transcribed under the "control" of the control sequences: i.e., the promoter will control the transcription of the mRNA derived from the coding sequence; and the ribosomes will bind at the ribosomal binding site to begin the translational process; and the stop codon used to terminate translation will be upstream from the transcriptional termination codon.

In addition to control sequences, it may be desirable to add regulatory sequences which allow for regulation of the expression of the human GPx gene relative to the growth of the host cell. Examples of regulatory systems are those which cause the expression of a gene to be turned on or off in response to a chemical or physical stimulus, including the presence of a regulatory compound. In prokaryotic systems these would include the lac and trp operator systems. Other examples of regulatory sequences are those which allow for gene amplification. In eukaryotic systems, these include the dihydrofolate reductase gene which is amplified in the presence of methotrexate, and the methallothionein genes, which are amplified with heavy metals. In these cases, the sequence encoding human GPx would be placed in tandem with the regulatory element.

Other types of regulatory elements may also be present in the vector, i.e. those which are not necessarily in tandem with the sequence encoding human GPx. An example is the SV_{40} enhancer sequence, which, by its mere presence, causes an enhancement of expression of genes distal to it.

Modification of the sequence encoding human GPx, prior to its insertion into the replicon, may be desirable or necessary, depending upon the expression system chosen. For example, in some cases, it may be necessary to modify the sequence so that it may be attached to the control sequences with the appropriate orientation, i.e. to maintain the reading frame. In some cases, it may be desirable to add sequences which cause the secretion of the polypeptide from the host organism, with subsequent cleavage of the secretory signal. In addition, in some cases, it may be desirable to alter the opal stop codon (reasons for this

alteration are discussed below). The techniques for modifying nucleotide sequences utilizing cloning are well known to those skilled in the art. They include, e.g., the use of restriction enzymes, of enzymes such as Bal31 to remove excess nucleotides, and of chemically synthesized oligonucleotides for use as adapters, to replace lost nucleotides, and in site directed mutagenesis. See, e.g., Maniatis, Fritsch & Sambrook, MOLECULAR CLONING: A LABORATORY MANUAL (1982); DNA CLONING, Volumes I and II (D.N. Glover, et al ed. 1985); NUCLEIC ACID HYBRIDIZATION (B.D. Hames & S.J. Higgins eds. 1984).

Modification of the sequence encoding human GPx is also necessary if polypeptides substantially similar to human GPx are to be synthesized. These polypeptides differ in some engineered way from the native enzyme. E.g., if a fragment of human GPx is the desired product, the sequence encoding the enzyme would be modified to remove the undesired sequences corresponding to the amino acids which are to be deleted. If an active fragment of human GPx is the desired product, the deleted sequences most likely would be in the regions of the amino- and/or the carboxy- terminus. However, the region encoding the selenocysteine moiety would be retained, as would other regions encoding portions of the enzyme which are necessary for activity.

It may also be of interest to synthesize analogs of human GPx and analogs of fragments of human GPx. Such analogs may vary in their substrate specificity. E.g., in pig heart there are apparently two forms of glutathione peroxidase, one of which catalyzes the reduction of H_2O_2 and lipid peroxides, and another which catalyzes the reduction of phospholipid peroxides and intact peroxidized

membranes. Despite the difference in substrate specificity, the mechanism of action of both forms of enzyme appears to be the same, and the amino acid compositions are similar. Nevertheless, there is evidence suggesting that there may be variations in the active site of the two analogous enzymes. F. Ursini et al. (1985), Biochim. Biophys. Acta 839, 62. Thus, it may be desirable to modify the sequence encoding the active site of human GPx so that the product polypeptide will preferentially reduce one type of substrate, e.g., phospholipid peroxides which are implicated in membrane damage.

It may also be of interest to synthesize analogs or fragments of human GPx which differ in their hydrophobicity, allowing greater or lesser interactions with membranes, or with liposomes. This may be accomplished by substituting hydrophobic amino acids for hydrophilic amino acids in some of the external domains of the polypeptide or vice versa. Such changes in hydrophobicity are accomplished by modifying the sequences encoding the specific amino acids which are to be substituted.

Polypeptides which are substantially similar to human GPx or to its fragments but which lack the active site may also be synthesized. In this case the sequence encoding the enzyme would be modified so that those codons encoding the amino acids of the active site would be deleted.

The above are examples of the way the human GPx polypeptide can be modified by modification of the sequence encoding human GPx. These examples are not meant to be exhaustive, and one skilled in the art can readily determine other modifications which would be useful. All of these modifications may be accomplished

using the above cited techniques and references concerning the modification of nucleotide sequences.

The appropriately modified sequence encoding a polypeptide substantially similar to human GPx, including human GPx, may be ligated to the control sequences prior to insertion into a vector. Alternatively, the coding sequence can be cloned directly into an expression vector which already contains the control sequences and an appropriate restriction site. For expression of the human GPx-type polypeptide in prokaryotes and in yeast, the control sequences will necessarily be heterologous to the coding sequence. In addition, for expression in prokaryotes, the coding sequence will be free of introns. In cases where human GPx is to be expressed in cell lines derived from vertebrates, the coding sequence can be with or without introns, and the control sequences may be either heterologous or homologous.

A number of replicons which may be used to construct prokaryotic expression vectors are known in the art. See, e.g., U.S. Patent Nos. 4,440,859; 4,436,815; 4,431740, 4,431739; 4,428941; 4,425,437; 4,418149; 4,422,994; 4,366, 246; 4,342, 832. Replicons which may be used to construct yeast expression vectors are also known in the art. See, e.g., U.S.Patent Nos. 4,446,235; 4,443,539; 4,430,428. An example of a replicon which can be used to construct an expression vector for mammalian host cells is described in commonly owned patent application no. 921,730, the disclosure of which is incorporated herein.

The choice of an expression system is dependent, in part, upon whether or not selenocysteine is to be incorporated into the polypeptide during translation. If the modified amino acid residue is to

be incorporated translationally, a host system which contains the system to carry out such translational incorporation is required. Examples of such systems include a variety of immortalized mammalian cells, including mouse cells, monkey cells, and cell lines derived from hamster cells (3T3, Vero, Chinese Hamster Ovary cells). Primary cell lines derived from vertebrates may also be used. It should be noted that in most cases these cell lines will contain directly the system for translating the opal stop codon, so that modification of the sequence encoding human GPx need not be altered to accommodate the translation.

An example of a prokaryotic expression host which can directly incorporate a selenocysteine during translation is E. coli. In these organisms selenocysteine is translationally incorporated into formate dehydrogenase. F. Zinoni et al. (1986). Proc. Natl. Acad Sci. USA 83, 4650. Moreover, the selenocysteine residue is incorporated in response to the opal nonsense codon (TGA).

In some cases, the host may contain the system for the translational incorporation of selenocysteine, but the triplet codon signaling the incorporation may not be the opal codon. If this type of host is used, the sequence encoding human GPx will have to be modified to replace the opal codon with the appropriate signal codon. The techniques for modification of the polynucleotide sequence have been discussed supra. In other cases, a host may be selected which is able to modify a pre-existing tRNA to Se-Cys tRNA.

If the selenocysteine residue is not to be incorporated translationally into the polypeptide, other expression hosts may be used. However, in using these systems the sequence encoding the opal stop codon may

have to be modified to allow translation to progress. In this case, the modification will encode for either cysteine or another amino acid(s) in place of selenocysteine.

Alternatively, the host will contain a suppressor system which allows translation of the opal stop codon. A variety of suppressor systems are known in the art. Cf. D.A. Steege and D.G. Soll (1979) in BIOLOGICAL REGULATION AND DEVELOPMENT OF GENE EXPRESSION (R.F. Goldberger, ed.) Vol 1,pp.433-475 (bacterial suppressor systems), F. Sherman (1981) in THE MOLECULAR BIOLOGY OF THE YEAST SACCHAROMYCES: LIFE CYCLE AND INHERITANCE (J.N. Strathern, E.W. Jones and J.R. Broach, eds.) pp.463-486 (yeast suppressor systems), and D. Hatfield, (1985) Trends in Biol. Sci. 10, 201-204 (suppressor systems in higher eukaryotes). These suppressor systems may be naturally occurring in the host, or may be incorporated into the host via vectors, either naturally occurring or engineered.

Selenium may be incorporated into human GPx post-translationally. One method for post-translational incorporation would be by synthesizing the polypeptide fragments which flank the selenocysteine molecule, and then joining in the proper orientation, the two polypeptides to the selenocysteine via peptide linkages. Another method would entail the incorporation of the selenium moiety into the appropriate site of a polypeptide analog of human GPx, which does not contain the selenium moiety.

Depending on the expression system and host selected, a polypeptide which is substantially similar to human GPx, including human GPx, or an analog of human GPx, or a fragment of human GPx, is produced by growing host cells transformed by an expression vector described

above under conditions whereby the polypeptide is expressed. The synthesized polypeptide is then isolated from the host cells and purified. If the expression system secretes the enzyme into growth media, the protein can be purified directly from the media. If the recombinant polypeptide is not secreted, it is isolated from cell lysates. The selection of the appropriate growth conditions and recovery methods are within the skill of the art.

Isolation of the newly synthesized polypeptide depends upon an assay system by which the polypeptide may be detected. These assay systems would be obvious to one skilled in the art. For example, if the newly synthesized polypeptide exhibits glutathione peroxidase activity, the polypeptide can be detected by assaying for that enzymic activity. Cf K. Takahashi et al. (1986) J. Clin. Invest. 77, 1402, for an assay for human GPx activity.

It is also possible to detect the newly synthesized polypeptide by immunoassay using antibodies to polypeptides substantially similar to human GPx, including human GPx. In this case, the type of antibody used in the assay will reflect the expected presence or absence of specific known epitopes. E.g., if the newly synthesized polypeptide is an analog of human GPx or a fragment of human GPx which lacks the seleno moiety, the preparations of antibodies used for detection would contain those directed towards parts of the molecule other than the epitope containing selenocysteine. On the other hand, if the newly synthesized polypeptide is human GPx including substantially similar polypeptides, or an active fragment thereof, all of which may be expected to contain the selenocysteine moiety, the preparations of antibodies used for immunoassay would

contain those directed to the epitope containing the selenocysteine moiety. The techniques of immunoassay are well known to those skilled in the art., and a method of preparing polyclonal antibodies to human GPx has been described. See K. Takahashi, J. Clin. Invest., supra.

In general, recombinant production of human GPx can provide compositions of that enzyme substantially free of contaminating proteins, i.e. of at least 95% purity. The ability to obtain high levels of purity is a result of recombinant expression systems which can produce human GPx in substantial quantities vis-a-vis in vivo sources. Thus, by applying conventional techniques to recombinant cultures, human GPx compositions of substantial purity and amount are obtainable.

It should be noted, that with the sequence data of the present invention, production of human GPx is not restricted to recombinant methods. It may also be synthesized by chemical methods such as solid phase peptide synthesis. Such methods are known to those skilled in the art.

The synthetic polypeptides which are substantially similar to human GPx, including human GPx, can be used to produce antibodies, both polyclonal and monoclonal. If polyclonal antibodies are desired, a selected mammal (e.g., mouse, rabbit, goat, horse, etc.) is immunized with purified human GPx polypeptide or fragment thereof, or analog thereof, or fragment of an analog thereof. Serum from the immunized animal is collected and treated according to known procedures. If serum containing polyclonal antibodies to human GPx contains antibodies to other antigens, the human GPx polyclonal antibodies can be purified by immunoaffinity chromatography. See, e.g., K. Takahashi, J. Clin.

Invest., supra., for a method of preparing polyclonal antibodies to human GPx.

Monoclonal antibodies to human GPx can also be readily produced by one skilled in the art. The general methodology for making monoclonal antibodies by hybridomas is well known. Immortal antibody-producing cell lines can be created by cell fusion, and also by other techniques such as direct transformation of B lymphocytes with oncogenic DNA, or transfection with Epstein-Barr virus. See, e.g., M. Schreier et al., HYBRIDOMA TECHNIQUES (1980); Hammerling et al., MONOCLONAL ANTIBODIES AND T-CELL HYBRIDOMAS (1981); Kennett et al., MONOCLONAL ANTIBODIES (1980); see also U.S. Patent Nos. 4,341,761; 4,399,121; 4,427,783; 4,444,887; 4,452,570; 4,466,917; 4,472,500, 4,491,632; and 4,493,890. Panels of monoclonal antibodies produced against human GPx can be screened for various properties; i.e., for isotype, epitope, affinity, etc. Monoclonal antibodies directed against specific epitopes are useful in defining interactions of human GPx with its various substrates. In addition, monoclonal antibodies are useful in purification, using immunoaffinity techniques, of native or recombinantly produced human GPx.

In the preparation of antibodies, prior to immunization, it may be desirable to increase the immunogenicity of human GPx, or particularly of its fragments. This can be accomplished in any one of several ways known to those of skill in the art. For example, the antigenic peptide or polypeptide may be administered linked to a carrier. For example, a fragment may be conjugated with a macromolecular carrier. Suitable carriers are typically large, slowly metabolized macromolecules such as: proteins;

polysaccharides, such as latex functionalized sepharose, agarose, cellulose, cellulose beads and the like; polymeric amino acids, such as polyglutamic acid, polylysine, and the like; amino acid copolymers; and inactive virus particles.. Especially useful protein substrates are serum albumins, keyhole limpet hemocyanin, immunoglobulin molecules, thyroglobulin, ovalbumin, and other proteins well known to those skilled in the art.

The protein substrates may be used in their native form or their functional group content may be modified by, for example, succinylation of lysine residues or reaction with Cys-thiolactone. A sulfhydryl group may also be incorporated into the carrier (or antigen) by, for example, reaction of amino functions with 2-iminothiolane or the N-hydroxysuccinimide ester of 3-(4-dithiopyridyl) proprionate. Suitable carriers may also be modified to incorporate spacer arms (such as hexamethylene diamine or other bifunctional molecules of similar size) for attachment of peptides.

If human GPx or polypeptides substantially similar to human GPx are to be used therapeutically, it may be desirable to link the polypeptide molecule to an efficient system to deliver the GPx to the appropriate site, and which will also protect the polypeptide from proteolysis, and at the same time cause a controlled delivery of the polypeptide. Systems for the delivery of molecules are known to those of skill in the art, and are reviewed, for e.g., in Poznansky, M.J., et al, in DRUG DELIVERY SYSTEMS (R.L. Juliano, ed, Oxford, N.Y., 1980), pp.253-315, and in Poznansky, M.L. (1984), Pharm Revs 36, 277.

An example of a delivery system which may be utilized for a polypeptide substantially similar to

human GPx is a liposome delivery system. Liposome delivery systems have been proposed for a variety of therapeutic products, particularly those which are administered parenterally. Liposomes have the potential of providing a controlled "depot" release of the administered GPx polypeptide over an extended time period, and of protecting the administered polypeptide from protease degradation. Liposomes can also alter the tissue distribution and uptake of the administered polypeptide. In addition, targeting macromolecules can be added to the liposome composition.

Several methods for preparing liposomes with entrapped compounds are known, and may be used for preparations of liposomes containing a polypeptide substantially similar to human GPx. For example, in one method, vesicle forming lipids are deposited as a thin film on the sides of a flask, and slowly rehydrated by addition of an aqueous buffer. The compound to be entrapped may be included either in the lipid film or in the aqueous hydration medium, depending upon the hydrophobicity and stability of the compound. The liposomes that form are multilamellar vesicles (MLVs) having heterogeneous sizes between about 0.05 and 10 microns.

The MLVs may be subsequently processed, typically by homogenization, sonication, or membrane extrusion, to produce smaller particles and a more uniformly sized suspension. Liposome sizing down to about 0.2-0.4 microns is generally preferred. Liposomes in this size range can be sterilized by passage through a 0.45 micron depth filter, have less tendency to aggregate, and also may show more favorable organ distribution when administered intravenously. Gabizon, A., et al (1982), Cancer Research 42, 4734.

Alternative methods for preparing liposomes are also known. E.g., they may be prepared by solvent injection, in which a lipid-in-solvent solution is injected into an aqueous medium, as described by, for example, Cafiso, D.S. (1981), Biochim. Biophys. Acta 649, 129. They can also be prepared in a reverse evaporation phase method, as described by Szoka, F. Jr., et al. (1980), Ann. Rev. Biophys. Eng. 9, 467. More specifically, a lipid-in solvent solution is mixed with an aqueous medium, and emulsified to form a water-in-oil emulsion. Removal of the lipid solvent produces a reverse-phase lipid gel which is then agitated, preferably in the presence of added aqueous medium, to form reverse-phase evaporation vesicles characterized by relatively large sizes and one to a few bilayer shells.

Since encapsulation is usually relatively poor with hydrophilic substances, it may be desirable to remove the free polypeptide from the composition containing the liposome encapsulated polypeptide. This can be achieved by treating the liposomes by known methods, for example, by molecular sieve chromatography, by centrifugation, or by diafiltration.

The purified human GPx compositions of the present invention are useful in several regards. First, they can be used to overcome genetic deficiencies in the patient's enzyme, particularly in individuals who would suffer from hemolytic episodes as a result of exposure to drugs or xenobiotics which generate superoxide radicals, H_2O_2, or lipid peroxides in red cells. They also have potential therapeutic use in overcoming the effects caused by ischemic trauma, as an adjunct to hypobaric oxygen treatment, as a treatment for cataracts, and in treatment of mental disorders related to aging, or which are currently treated with

chlorpromazine. Human GPx may also have therapeutic use in preventing the development of coronary heart disease, which is related to its affect on prostacyclin synthesis, as well as the deleterious effects of the deficiency leading to premature red cell aging and damage to the endothelial lining of blood vessels.

Human GPx will also be useful as a tool in designing activators and inhibitors of the enzyme. These activators and inhibitors have several uses. E.g., they would be useful in delineating the mechanism by which human GPx acts with its several types of substrates, eg. lipid peroxides, peroxidized DNA, H_2O_2, and the intermediates in the arachidonic cascade. In addition, therapeutically these enzyme modulators could be useful in modulating the inflammatory and hypersensitivity reactions which result from products produced by the arachidonic acid cascade.

The sequence encoding human GPx may also be inserted into vectors which are useful for human gene therapy, and used as a treatment for individuals who are genetically deficient in the enzyme. These vectors carry the information which allow the insertion of the sequence into the treated individuals genome. The vectors may be either of the DNA type, e.g. SV_{40} derivatives, or of the RNA type, e.g. the RNA retroviruses. Methods of inserting sequences into these oncoviruses are known to those skilled in the art.

Recombinant polypeptides substantially similar to human GPx, including human GPx and fragments thereof, are also useful for the synthesis of antibodies to these polypeptides; the methods for production of the antibodies has been discussed above.

The antibodies to polypeptides substantially similar to human GPx, including human GPx, would have

many uses; e.g., such antibodies, both polyclonal and monoclonal, would find use in the purification of human GPx by immunoaffinity chromatography.

The antibodies would also be useful for the diagnosis of glutathione peroxidase deficiency. Moreover, since selenium deficiency causes a reduction in human GPx levels (K. Takahashi et al., J. Clin. Invest., supra.,) the antibodies would also be useful for diagnosing selenium deficiency in humans.

In addition, the human GPx fragments and analog polypeptides as well as monoclonal antibodies to specific epitopes on glutathione peroxidase will be useful in determining how the enzyme interacts with its various substrates, as well as how the enzyme may be targeted for hydrophilic and hydrophobic sites. This knowledge will be useful in designing analogs of human GPx for specific therapeutic uses. E.g., analogs may be designed which will affect the arachidonic cascade by virtue of limiting activity with substrates. Other types of analogs may be designed which are highly suitable for preventing membrane damage which is associated with lipid peroxidation.

Described below are examples of the present invention which are provided only for illustrative purposes, and not to limit the scope of the present invention. In light of the present disclosure, numerous embodiments within the scope of the claims will be apparent to those of ordinary skill in the art.

Experimental

Sequences encoding human GPx DNA, derived from cDNA, were cloned according to the following procedure.

1. Materials and Methods

Enzymes were purchased from commercial sources, and used according to the manufacturers' directions. Radionucleotides and nitrocellulose filters were also purchased from commercial sources.

Plasmids were propagated in E. coli strain HB101. Lambda (λ) bacteriophage gt10 was propagated in E. coli strain C600ΔHFL.

DNA was sequenced by the dideoxynucleotide chain termination method in bacteriophage m13. Sanger et al. (1977), Proc. Natl. Acad. Sci. USA 74, 5463. To overcome problems with band compression, which are sometimes observed in GC rich regions, 7-deazoguanosine was used. Barr, P., et al (1986) Biotechniques 4, 428-432.

DNA was synthesized by automated oligomer synthesis on a glass support using an automated DNA synthesizer (B.W. Warner et al. (1984) DNA 3, 401-411). and nucleoside N,N-diisopropyl phosphoramidites (L. McBride and M. Caruthers (1983) Tetrahedron Lett. 24, 245. And S.P. Adams, et al. (1983) J. Am. Chem. Soc. 105, 661-663).

In the cloning of DNA fragments, all DNA manipulations were done according to standard procedures. See, Maniatis et al., MOLECULAR CLONING, supra. Restriction enzymes, T_4 DNA ligase, DNA polymerase I, and other biological reagents were purchased from commercial suppliers and used according to the manufacturers' directions. Double-stranded DNA fragments were separated on agarose gels and isolated by electroelution.

cDNA libraries were prepared by standard techniques in λgt10. See, T.V. Huynh et al., in DNA CLONING: VOL/1. (Glover, Ed.), supra. Double stranded

cDNA was prepared essentially as described. U. Gubler and B.J. Hoffman (1982), supra. After methylation of internal EcoRI sites and the addition of EcoRI linkers, the cDNA was ligated into λgt10. T.V. Huynh et al. (1985) in DNA CLONING (D.M. Glover, ed) pp 49-78. The phage were packaged and recombinants selected by plating on E. coli strain BNN102.

Nucleotide homologies between the cloned sequences encoding human, bovine, and mouse GPx were deduced using a computer alignment program. This program was also utilized to test for the presence of significant homology of these cloned sequences with sequences in the data base "organelles" wherein the sequences of the entire mitochondrial bovine and genome reside. Homologies of the proteins deduced from the cloned DNA sequences were deduced using a computer alignment program.

2. Cloning of cDNA Sequences Encoding Bovine GPx

A cDNA library derived from bovine pituitary was prepared. Recombinants carrying the gene encoding bovine GPx were selected by probing with two specially designed synthetic oligomers shown in Fig. 1.

These oligomers were designed from the known amino acid sequence of bovine GPx. W. A. Gunzler et al. (1984), See prev. so as to anneal to regions of the coding strand which possess low degeneracy in their code, namely regions coding amino acids 108 through 125 as numbered in Figure 3 (for the 53-mer), and 81 through 95 (for the 44-mer). Where degeneracies were encountered, the following prioritized principles were exercised in selecting a nucleotide: (1) consistency with the codon usage bias observed in bovine DNA was maintained; (2) in cases where potential mismatches

might arise, GT pairs rather than other mismatches were favored; and (3) the sequence 5'-CG-3', seldom observed in the coding strand of eukaryotes, was also presumed to occur infrequently, and was thus avoided.

To select the recombinants carrying the bovine GPx sequences, the culture plates were replica plated in duplicate onto nitrocellulose filters. The filters were probed by hybridization with the oligomers which had been terminally labeled with ^{32}P, followed by washing in 4XSSC, 0.1% SDS at 50°C, for 60 min. The conditions for hybridization and for the terminal labelling of the probe are given in T. Maniatis et al., Molecular Cloning Supra. Approximately 20 plaques gave strong signals; 10 of these were isolated and plaque purifed by replating on E. coli strain BNN102 and sequencing of the DNA insert.

The EcoRI insert DNA encoding bovine GPx from the 10 clones was subcloned into pBR322, using standard techniques. For DNA sequencing, the inserts were reexcised from the pBR322 with EcoRI. The sequence of the largest insert, 830 bp, is presented in Fig. 2.

2. Cloning of cDNA Sequences Encoding Human GPx

A cDNA library derived from human kidney was prepared. Recombinants carrying sequences encoding human GPx were selected by using a probe derived from the 830 bp sequence encoding bovine GPx.

More specifically, the 830 bp fragment encoding bovine GPx was excised from pBR322 with EcoRI, and isolated by gel electrophoresis. The fragment was radiolabelled by nick translation. See T. Maniatis et al., MOLECULAR CLONING, supra. Recombinants were selected by replica plating and lifting DNA onto nitrocellulose in duplicate, followed by probing the

filters by hybridization with the radiolabelled probe, and washing the filters in 0.1XSSC, 0.1% SDS at 25°C for 60 min. To isolate clones, phage from 10 plaques were plaque purified, and the EcoRI insert DNA was sequenced. Three clones encoding fragments of human GPx were obtained: a clone containing a 550 bp insert (hGPx1) corresponding to the 3' end of the GPx mRNA; a clone containing a 280 bp insert (hGPx2) corresponding to the 5' end of the GPx mRNA; and a clone containing a 560 bp insert (hGPx3). The 560 bp clone was found to correspond to a portion of 3' end of the GPx mRNA and also contained a 278 bp intron.

The EcoRI insert DNAs encoding human GPx from these clones was subcloned into the EcoRI site of pBR322, using standard techniques. For DNA sequencing, the inserts were reexcised from the pBR322 with EcoRI.

4. Cloning of cDNA Sequences Encoding Mouse GPx

Utilizing a cDNA library prepared from mouse placenta, clones carrying sequences encoding mouse GPx were isolated by the procedures described for the isolation of clones carrying sequences encoding human GPx, described above. Using these procedures, two clones were isolated; one clone contained a fragment of about 550 bp, and the other clone contained a fragment of about 300 bp. These inserts were subcloned into pBR322, and their DNA inserts sequenced after excision with EcoRI.

5. Homologies of the cDNA Inserts Encoding Human, Mouse, and Bovine GPx.

The nucleotide sequence encoding human GPx is shown in Fig. 2, which compares the human sequence with the bovine sequence and a mouse sequence. The cDNAs

used in deriving the sequence were those described above, i.e., the two human sequences which did not contain the intron, the two mouse sequences, and the full length bovine sequence. The sequences were derived from sequencing both strands of the cDNAs. However, the sequencing was not done across the internal EcoRI sites of the human and mouse DNA. Nevertheless, the homology with bovine sequences across these restriction sites and their flanking regions is high and contiguous, indicating that the mouse and human fragments are properly aligned.

In Fig. 2, spaces indicated by hyphens have been introduced to maximize homology between the DNAs. Positions of identity are indicated by asterisks. The features of the sequences are highlighted as follows. Enclosed within boxes are the presumed translational start codon, ATG; the opal codon within the structural gene, TGA; and the codon used to stop translation. Putative polyadenylation signals are double-underlined. The location of the intron in the human and the mouse sequences is shown as a vertical broken line and marked IVS. Nonoverlapping inverted repeats greater than 6bp which are separated by less than 30 bp are shown by arrows.

A comparison of the human, mouse, and bovine cDNA sequences shows the following. The GC contents of the human, mouse, and bovine cDNA are 62%, 62%, and 59%, respectively. In those codons where wobble with A or T is permitted in the third position, G or C is favored in that position 81%, 73%, and 82%, for human, mouse and bovine, respectively. Within the coding region, the nucleic acid homology for all three species is approximately 83%, although bovine carries a 15 bp insert (Fig. 2). The homologies within the other

regions reflect more variation. Within the 3'-untranslated region the homology is 71% (bovine/human), 69% (mouse/human), and 67% (bovine/mouse). In the 5'-untranslated region, homology between bovine and mouse is approximately 35%. The human and mouse cDNAs possess a canonical polyadenylation signal (AATAAA) at positions 799 and 814, respectively. It is presumed that the bovine cDNA clone did not reveal this feature because of its relative shortness.

6. Localization of the genes encoding GPx

The cellular location of the gene coding for GPx is of interest because of the unusual coding event which occurs, TGA-SeCys. Since GPx has two cellular locations, i.e. mitochondrial and extramitochondrial, it is important to determine whether the genes encoding the cDNAs isolated are located in the mitochondrion or in the nucleus. The cellular location of the genes was established by a computer search of the total bovine and human mitochondrial genome libraries using sequences homologous to the isolated bovine and human GPx cDNAs. The lack of appreciable homology of the isolated cDNAs to the mitochondrial libraries indicates that the genes encoding these GPxs are of nuclear origin.

7. Analysis of the Deduced Amino Acid Sequences of Human, Mouse, and Bovine GPxs

The amino acid sequence for human GPx, as deduced from the nucleotide sequence for cDNA encoding the polypeptide is shown in Fig. 3, which also shows the homologies of the mouse and bovine sequences to that of the human. In the Figure, maximum homology is achieved by including spaces, which are indicated by dashes. The

symbol # indicates inferred selenocysteine residues (discussed below). The terminal amino acids of the mature bovine protein as deduced by chemical methods (W. A. Gunzler, et al. (1984) (see page 3)) and by X-ray diffraction analysis (O. Epp, et al. (1983) Eur J. Biochem. 133, 51-69) are shown by closed and open triangles, respectively. Potential glycosylation sites are indicated by underlined amino acids. The homologous location of the mouse and human introns is shown by an arrow (IVS). Amino acids which are proposed to bind glutathione or play a functional role during catalysis are shown by circles. The sequences are shown in single-letter code and correspond to the three letter code as follows: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.

The opal stop codon, TGA, within the cDNA sequence encoding human GPx encodes the unusual amino acid residue, selenocysteine. This is inferred from a comparison of the bovine cDNA sequence with the derived amino acid sequence (Figs. 2 and 3) and the published amino acid sequence (W. A. Gunzler, et al., (1984)(see page 3)), and the homologies between the human and bovine sequences. A similar conclusion has been reached for mouse GPx. I. Chambers et al (1986), EMBO J. 5, 1221.

Based upon the deduced sequences, preprocessed human GPx contains 201 amino acids. In comparison, preprocessed bovine and mouse GPx contain 205 and 201 amino acids, respectively (Fig. 3). Approximately 87% of all human residues are homologous with the bovine residues, although bovine GPx possesses a 5 amino acid insert. Homology between human and mouse residues is

85%. An additional 7 to 10%, not exhibiting identity between pairs of species, arise from conservative replacements in the genetic code. It has been proposed that three bovine residues, Arg-57, Met-147 and Arg 184, bind glutathione during catalysis and that Trp-165 is involved catalytically. These residues are apparently conserved in both human and mouse GPx.

Certain structural features of human GPx may also be inferred from the cDNA sequences and the deduced amino acid sequences. First, nucleotide sequencing of one of the human GPx cDNA EcoRI inserts showed the presence of a 278 bp intron which exists between the nucleotides coding for Gln-82 and Glu-83 (Figs. 2 and 3). This position is homologous to a 216 bp intron in mouse. Chambers et al (1986), EMBO J., supra. A proposed function of introns is to separate two protein domains. In the case of human GPx, the intron would separate the α_1-helix of the polypeptide from the α_2-helix and the remainder of the polypeptide.

The analysis of the deduced amino acid sequence also shows features which are relevant to processing of the nascent human GPx. First, the N-terminus of the nascent polypeptide lacks potential dibasic processing signals, suggesting a lack of a specific cleavage event. In addition, the classical features of a potential N-terminal secretory signal-sequence with its accompanying cleavage site are also lacking. Thus, it would appear that human GPx does not enter the secretory pathway. Finally there is a lack of potential N-glycosylation sites (Asn-X-Ser/Thr), suggesting that the polypeptide is not N-glycosylated.

Analysis of the deduced amino acid sequences of bovine and mouse GPxs shows that the above also pertains to those enzymes, with one exception. Both the bovine

and mouse amino acid sequences contain potential N-glycosylation sites. However, because the proteins are otherwise highly homologous, and because comparison of the locations of the potential glycosylation sites shows a marked variance in their positions, it suggests that the mouse and bovine enzymes also are not N-glycosylated.

HB101(phGPx1), and HB101(phGPx2) have been deposited with the American Type Culture Collection (ATCC), 12301 Parklawn Dr., Rockville, Maryland 20852, and have been assigned Accession Nos._____, and _____, respectively. HB101(phGPx1) contains the 550 bp insert encoding a portion of human GPx. HB101(phGPx2) contains the 200 bp insert encoding a portion of human GPx. These deposits are intended for convenience only, and are not required to practice the present invention in view of the description herein. The human GPx DNA sequences in the deposited material are incorporated herein by reference.

While the present invention has been illustrated above by certain specific embodiments, it is not intended that these specific examples limit the scope of the invention as described in the appended claims.

CLAIMS

1. A polynucleotide construct encoding a polypeptide substantially similar to human glutathione peroxidase (GPx), said polynucleotide construct comprising a heterologous region, said heterologous region comprising a coding sequence for said polypeptide substantially similar to human GPx, wherein the polypeptide substantially similar to human GPx encoded within the heterologous region is selected from a group consisting of human GPx, a fragment of human GPx, an analog of human GPx, and an analog of a fragment of human GPx.

2. The polynucleotide construct of claim 1, wherein the polynucleotide is double stranded DNA.

3. The DNA construct of claim 2 wherein the coding sequence is for human GPx.

4. The DNA construct of claim 3 wherein the coding sequence is that shown in Figure 2.

5. The polynucleotide construct of claim 1, wherein the construct is further comprised of sequences necessary for replication of the construct.

6. The polynucleotide construct of claim 5, wherein the construct is further comprised of sequences allowing for expression of the coding sequence, said sequences comprising:

a promoter sequence;

a ribosomal binding sequence;

a sequence which causes the cessation of translation; and

a sequence which causes the cessation of transcription.

7. The polynucleotide construct of claim 6, wherein the sequences allowing expression of the coding sequence allow expression in eukaryotic cells.

8. The polynucleotide construct of claim 7, wherein the sequences allowing expression of the coding sequence allow expression in mammalian cells.

9. The polynucleotide construct of claim 7, wherein the sequences allowing expression of the coding sequence allow expression in yeast cells.

10. The polynucleotide construct of claim 6, wherein the sequences allowing expression of the coding sequence allow expression in prokaryotic cells.

11. An expression host harboring the polynucleotide construct of claim 6.

12. A composition comprising cDNA encoding human GPx.

13. A method of producing a recombinant polypeptide substantially similar to human GPx, wherein polypeptides substantially similar to human GPx is selected from a group consisting of human GPx, a fragment of human GPx, an analog of human GPx, and an analog of a fragment of human GPx, the method comprising:

a. providing a population of transformed cells containing a vector, said vector comprising

(1) a coding sequence for said polypeptide substantially similar to human GPx, and

(2) sequences allowing expression of said coding sequence in said cells, said sequences allowing expression comprising a promoter which controls transcription of said coding sequence, a ribosomal binding sequence, a translation termination sequence, and a transcription termination sequence,

b. growing said population of transformed cells under conditions whereby said polypeptide substantially similar to human GPx is expressed; and

c. recovering said polypeptide substantially similar to human GPx.

14. The method of claim 13 wherein said cells are prokaryotic.

15. The method of claim 13 wherein said cells are eukaryotic.

16. The method of claim 15 wherein the eukaryotic cells are yeast cells.

17. The method of claim 15 wherein the eukaryotic cells are mammalian cells.

18. A composition comprising a polypeptide synthesized by expression of the polynucleotide construct of claim 1.

19. A composition according to claim 18, further comprising a macromolecular structure with which the polypeptide is associated, wherein said macromolecular structure is a tissue specific targeting agent for the polypeptide.

20. A composition according to claim 18, further comprising a structure in which the polypeptide is encapsulated.

21. A composition according to claim 20, wherein the structure is a liposome.

22. A composition according to claim 20, further comprising a macromolecular structure with which the polypeptide is associated, wherein said macromolecular structure is a tissue specific targeting agent for said polypeptide.

5' GCCTTCTCCCCATTCACCTCACACTTCTCAAACAGCAT
AAAGTTAGGCTCAAA 3'
(53-mer)

3' ACGTTGGTCAAACCGGTGGTCCTTTTGCGGTTTTTGCTTCTTTA 5'
(44-mer)

Figure 1

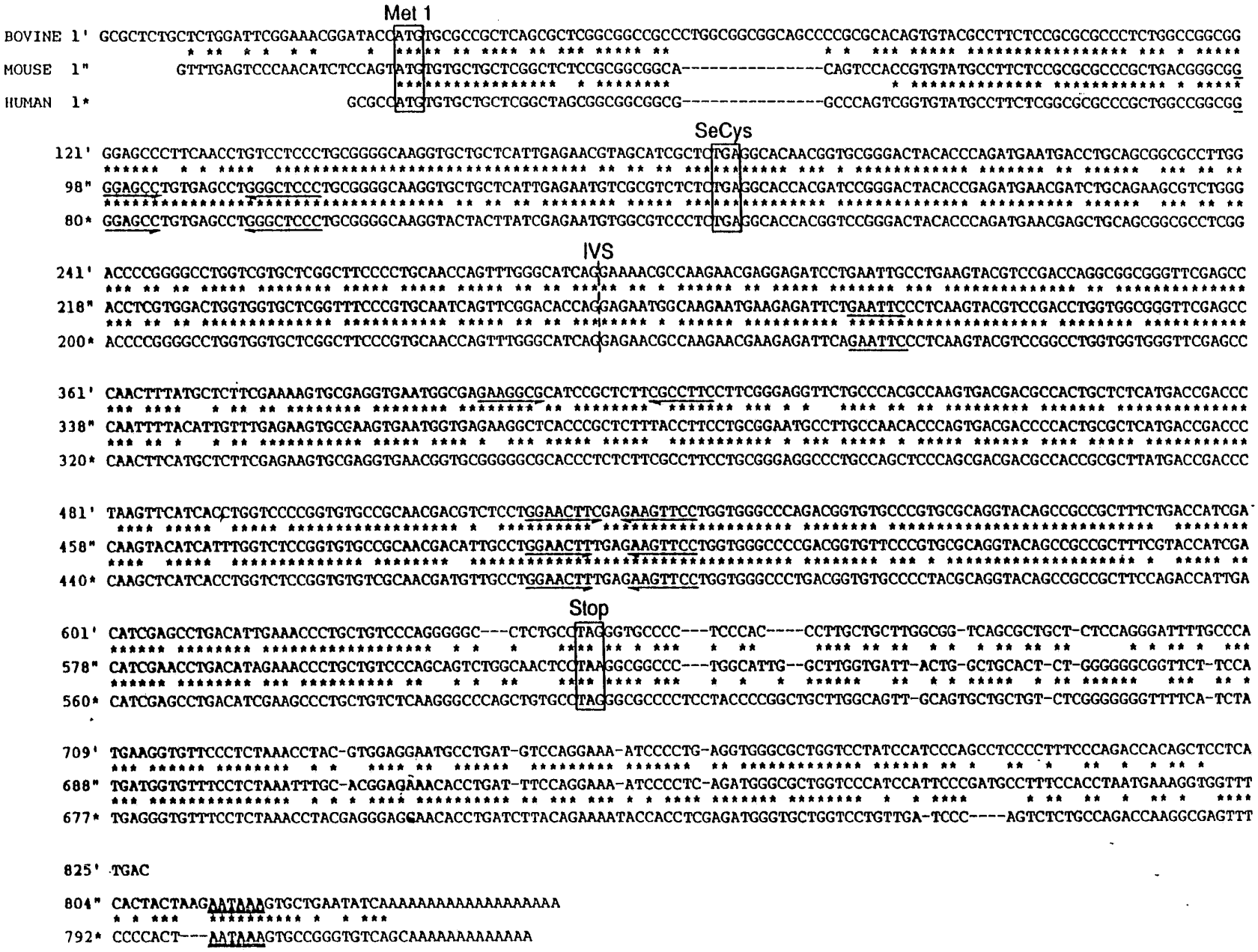

FIGURE 2

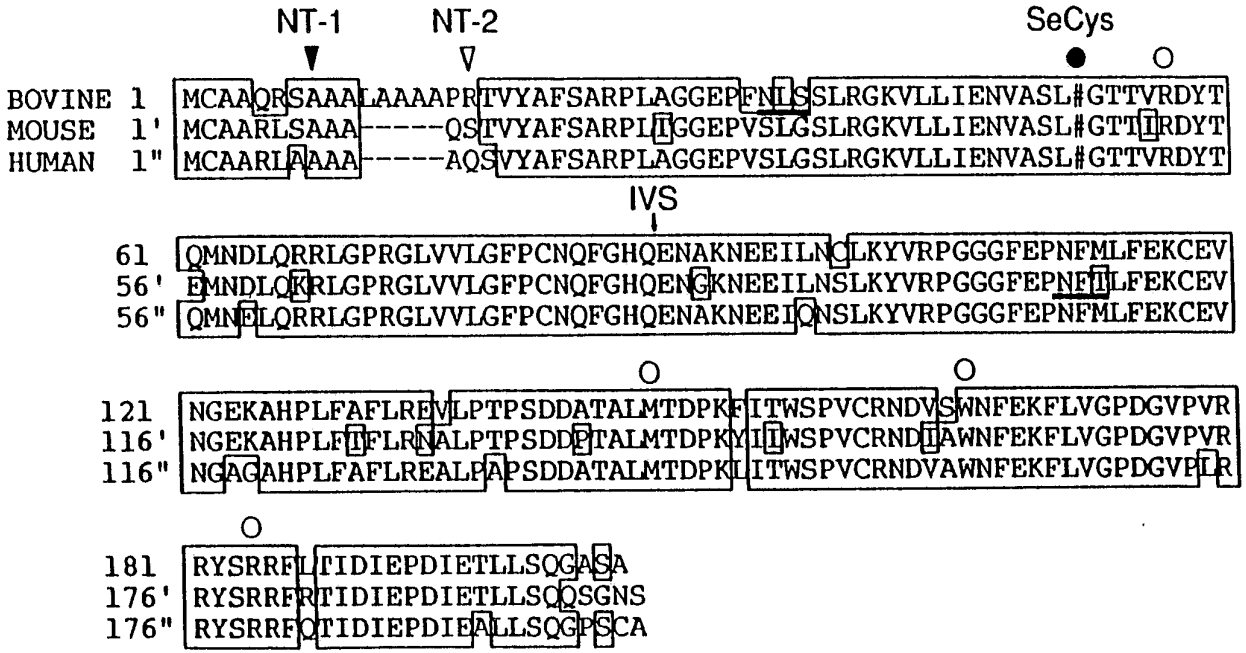
NT-1
NT-2
SeCys
BOVINE 1 MCAAQRSAAALAAAAPRTVYAFSARPLAGGEPFNLSSLRGKVLLIENVASL#GTTVRDYT
MOUSE 1' MCAARLSAAA-----QSTVYAFSARPLIGGEPVSLGSLRGKVLLIENVASL#GTTIRDYT
HUMAN 1" MCAARLAAAA-----AQSVYAFSARPLAGGEPVSLGSLRGKVLLIENVASL#GTTVRDYT
IVS
61 QMNDLQRRLGPRGLVVLGFPCNQFGHQENAKNEEILNCLKYVRPGGGFEPNFMLFEKCEV
56' EMNDLQKRLGPRGLVVLGFPCNQFGHQENGKNEEILNSLKYVRPGGGFEPNFILFEKCEV
56" QMNELQRRLGPRGLVVLGFPCNQFGHQENAKNEEIQNSLKYVRPGGGFEPNFMLFEKCEV
121 NGEKAHPLFAFLREVLPTPSDDATALMTDPKFITWSPVCRNDVSWNFEKFLVGPDGVPVR
116' NGEKAHPLFTFLRNALPTPSDDPTALMTDPKYIIWSPVCRNDIAWNFEKFLVGPDGVPVR
116" NGAGAHPLFAFLREALPAPSDDATALMTDPKLITWSPVCRNDVAWNFEKFLVGPDGVPLR
181 RYSRRFLTIDIEPDIETLLSQGASA
176' RYSRRFRTIDIEPDIETLLSQQSGNS
176" RYSRRFQTIDIEPDIEALLSQGPSCA

FIGURE 3

International Application No: PCT/ US88 / 00835

MICROORGANISMS

Optional Sheet in connection with the microorganism referred to on page _____, line _____ of the description [1]

A. IDENTIFICATION OF DEPOSIT [2]

Further deposits are identified on an additional sheet ☐ [3]

Name of depositary institution [4]

American Type Culture Collection

Address of depositary institution (including postal code and country) [4]

12301 Parklawn Drive
Rockville, Maryland 20852
United States of America

Date of deposit [5]	Accession Number [6]
3 April 1987	67381 67382

B. ADDITIONAL INDICATIONS [7] (leave blank if not applicable). This information is continued on a separate attached sheet ☐

C. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE [8] (if the indications are not for all designated States)

Japan
EPO

D. SEPARATE FURNISHING OF INDICATIONS [9] (leave blank if not applicable)

The indications listed below will be submitted to the International Bureau later [9] (Specify the general nature of the indications e.g., "Accession Number of Deposit")

E. ☐ This sheet was received with the international application when filed (to be checked by the receiving Office)

(Authorized Officer)

☐ The date of receipt (from the applicant) by the International Bureau [10]

was 07 JUN 1988

(Authorized Officer)

(January 1985)

Form PCT/RO/134 (January 1981)

INTERNATIONAL SEARCH REPORT

International Application No. PCT/US88/00835

I. CLASSIFICATION OF SUBJECT MATTER (if several classification symbols apply, indicate all) [6]

According to International Patent Classification (IPC) or to both National Classification and IPC

IPC(4): C07H 21/00; C12N 15/00; C12N 9/08; C12N 5/00 ; C12N 1/20 C12N 1/16
U.S. 536/27; 435/192, 240.2,253,255,172.3,320

II. FIELDS SEARCHED

Minimum Documentation Searched [7]

Classification System	Classification Symbols
U.S.	435/91,172.3,192,240.2,253,255,320 536/27, 935/10,14, 28,29,32,41,60,70,71,72

Documentation Searched other than Minimum Documentation
to the Extent that such Documents are Included in the Fields Searched [8]

Chemical Abstracts Data Base 1967-1988; BIOSIS DATA BASE 1969-1988.
Keywords RN 9013-66-5, Glutathione (W) Peroxidase,cloning, plasmid, vector, sequence see attachment.

III. DOCUMENTS CONSIDERED TO BE RELEVANT [9]

Category *	Citation of Document, [11] with indication, where appropriate, of the relevant passages [12]	Relevant to Claim No. [13]
X Y,P	NUCLEIC ACIDS RESEARCH, Volume 15, Number 13, issued 1987, July (Oxford, England) (G. Mullenbach et al)), "Sequence of a cDNA coding for human glutathione peroxidase confirms TGA encodes active site selenocysteine", see page 5484.	1-5 and 12 6-11 and 13-17
X Y,P	NUCLEIC ACIDS RESEARCH, Volume 15, Number 17, issued 1987, September (Oxford, England) (Y.Sukenaga et al), "cDNA sequence coding for human glutathione peroxidase", see page 7178.	1-5 and 12 6-11 and 13-17
X Y,P	NUCLEIC ACIDS RESEARCH , Volume 15, Number 23, issued 1987, December (Oxford, England) (K. Ishida et al), "Nucleotide sequence of a human gene for glutathione peroxidase", see page 10051.	1-5 and 12 6-11 and 13-17

* Special categories of cited documents: [10]

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

IV. CERTIFICATION

Date of the Actual Completion of the International Search	Date of Mailing of this International Search Report
14 JUNE 1988	27 JUN 1988
International Searching Authority	Signature of Authorized Officer
ISA/US	JAYME A. HULEATT

Form PCT/ISA/210 (second sheet) (Rev.11-87)

FURTHER INFORMATION CONTINUED FROM THE SECOND SHEET

V. ☐ OBSERVATIONS WHERE CERTAIN CLAIMS WERE FOUND UNSEARCHABLE [1]

This international search report has not been established in respect of certain claims under Article 17(2) (a) for the following reasons:

1. ☐ Claim numbers . because they relate to subject matter [2] not required to be searched by this Authority, namely:

2. ☐ Claim numbers , because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out [3], specifically:

3. ☐ Claim numbers_____, because they are dependent claims not drafted in accordance with the second and third sentences of PCT Rule 6.4(a).

VI. ☒ OBSERVATIONS WHERE UNITY OF INVENTION IS LACKING [2]

This International Searching Authority found multiple inventions in this international application as follows:

See Attachment.

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims of the international application.

2. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims of the international application for which fees were paid, specifically claims:

3. ☒ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claim numbers:

1-17

4. ☐ As all searchable claims could be searched without effort justifying an additional fee, the International Searching Authority did not invite payment of any additional fee.

Remark on Protest

☐ The additional search fees were accompanied by applicant's protest.

☐ No protest accompanied the payment of additional search fees.

III. DOCUMENTS CONSIDERED TO BE RELEVANT (CONTINUED FROM THE SECOND SHEET)

Category *	Citation of Document, with indication, where appropriate, of the relevant passages	Relevant to Claim No
Y	THE EMBO JOURNAL, Volume 5, Number 6, issued 1986, June (Oxford, England) (I. Chambers et al), "The structure of the mouse glutathione peroxidase gene: the selenocysteine in the active site is encoded by the termination codon TGA", see pages 1221-1227, see particularly pages 1221, 1222 and 1224-1226.	1-17
Y	CHEMICAL ABSTRACTS, Volume 100, Number 19, issued 1984, May 7 (Columbus, Ohio, U.S.A.)(W.A. Guenzler et al), "The amino acid sequence of bovine glutathione peroxidase, see page 239, column 2, the abstract No. 153063f. Hoppe-Seyler's Z. Physiol. Chem. 1984, 365(2), 195-212 (Eng.)"	1-17
A,P	TRENDS OF BIOCHEMICAL SCIENCES Volume 12, issued 1987, July (Cambridge, England) (I. Chambers et al) "A new puzzle is selenoprotein biosynthesis: selenocysteine seems to be encoded by the stop codon, UGA", see pages 255-256.	1-17

Attachment to PCT/ISA/210
Part VI.

Group I, claims 1-17 drawn to a polynucleotide construct encoding human glutathione peroxidase (GPx), an expression host, a cDNA encoding human GPx and a method of cloning GPx using the polynucleotide construct classified in Class 536, subclass 27, and Class 435, subclasses 172.3, 192, 240.2, 253, 255 and 320.

Group II, claims 18-22, drawn to GPx and conjugates and liposomes containing GPx.

PCT/US88/00835

Attachment to PCT/ISA/210
part II. Fields of Searched

DNA SEQUENCE DATA BASE SEARCH for the Human Glutathione peroxidase gene on Intelligenetics.